

# Application of Bayesian methods to modelling of extreme precipitation events

Kári Hlynsson

## 1 Introduction

Extreme precipitation events, such as flooding, pose significant risks to both infrastructure and human safety. Events are unpredictable, localised, and difficult to estimate due to sparse data, which are sourced from gauging stations built in close proximity to catchments such as streams, wells, lakes, canals, and reservoirs. [1] Accurate modelling and prediction is vital for the implementation of preventive measures to mitigate damage caused by such events, as well as serving as interesting applications a host of statistical methods explored here.

In this article, the application of Bayesian modelling methods, particularly Markov-chain Monte Carlo sampling, to annual maximum precipitation data totalling  $N = 132$  measurements is considered. The generalised extreme value (GEV) distribution is used to model the response layer with (i) fixed parameters, or a "stationary" model, and (ii) an evolving location parameter with respect to time, a "non-stationary" model. For the latter type, two models are proposed; the former implements a linear time trend, whereas the latter model allows a more flexible trend. Two information criteria, WAIC and LOOIC, are used for model comparison and selection. Posterior predictive checks are used for model evaluation. Data processing and visualisation was performed in the R language and MCMC sampling was implemented in Stan. [2, 3] Code and data used in the project are available at [https://github.com/lvthnn/gev\\_precipitation\\_data](https://github.com/lvthnn/gev_precipitation_data).

## 2 Preliminaries

### 2.1 Theoretical background

The generalised extreme value (GEV) distribution has parameters  $\mu \in \mathbf{R}$ ,  $\sigma \in \mathbf{R}_+$ , and  $\xi \in \mathbf{R}$ , termed the location, shape, and scale parameters, respectively. The cumulative density is

$$\text{GEV}(y | \mu, \sigma, \xi) = \begin{cases} \exp\left(-[1 + \xi(y - \mu)/\sigma]_+^{-1/\xi}\right), & \xi \neq 0, \\ \exp(-\exp[-(y - \mu)/\sigma]), & \xi = 0 \end{cases}$$

and support  $(-\infty, \mu - \sigma/\xi]$ ,  $(-\infty, \infty)$ , and  $[\mu - \sigma/\xi, \infty)$  for  $\xi < 0$ ,  $\xi = 0$ , and  $\xi > 0$ , respectively. This distribution arises naturally in a statistical setting involving normalised block maxima of random variables  $(X_n)_{n \in \mathbf{N}}$ , defined by  $(M_n - a_n)/b_n$  for suitable constants

$a_n \in \mathbf{R}_+$  and  $b_n \in \mathbf{R}$  where  $M_n = \max\{X_1, \dots, X_n\}$ , as by the Fisher-Tippett theorem, the only limiting distribution for  $M_n$  is the GEV distribution. [4, 5] In other words, if

$$\Pr\{(M_n - b_n)/a_n \leq y\} \rightarrow G(y),$$

then necessarily  $G(y) = \text{GEV}(y | \mu, \sigma, \xi)$  for suitable  $\mu \in \mathbf{R}$ ,  $\sigma \in \mathbf{R}_+$ , and  $\xi \in \mathbf{R}$ . If  $Y$  is a GEV random variable, then it has expectation

$$E[Y] = \begin{cases} \mu + \sigma[\Gamma(1 - \xi) - 1]/\xi & \xi \in (-\infty, 1) \setminus \{0\}, \\ \mu + \sigma\gamma & \xi = 0, \\ \infty & \xi \geq [1, \infty), \end{cases}$$

and variance given by

$$\text{var}[Y] = \begin{cases} \sigma^2[\Gamma(1 - 2\xi) - \Gamma(1 - \xi)^2]/\xi^2 & \xi \in (-\infty, 1/2) \setminus \{0\}, \\ \sigma^2\pi^2/6, & \xi = 0, \\ \infty & \xi \geq 1/2, \end{cases}$$

where  $\Gamma(\cdot)$  is the gamma function and  $\gamma \approx 0.5772$  is Euler's constant.

## 2.2 Converge diagnostics, model selection, and evaluation

For each model, four separate chains are run for a total of 9000 iterations and the first 1000 are discarded as burn-in. To ensure that the chains are sufficiently close to stationarity, we use trace and autocorrelation plots for visual confirmation of convergence, and the potential scale reduction factor, denoted  $\hat{R}$ , as well effective sample size  $\bar{\text{ESS}}$ . The former diagnostic compares parameter estimates between and within the different chains respectively. If the estimates are very similar,  $\hat{R}$  is close to one. Higher values may indicate that chains run simultaneously from different initial values have not converged and may thus indicate a problem with the sampler.

The latter quantity, the effective sample size, or ESS, is defined as

$$\text{ESS} = \frac{N_{\text{iter}}}{1 + 2 \sum_{l=1}^{\infty} \rho(l)},$$

where  $N_{\text{iter}}$  is the number of samples drawn, and  $\rho(l)$  is the autocorrelation between samples  $l$  iterations apart. Low effective sample size indicates that samples produced by the MCMC algorithm are highly correlated and thus provide a biased view of the posterior distribution. [6] Resultingly, parameter inference is unreliable.

WAIC and LOOIC are used to compare models and to select the optimal value of the hyper-parameter  $m$  in the case of piecewise hierarchical models. Both are asymptotically equivalent to cross validation, although LOOIC has been shown to outperform in WAIC in certain cases, such as uninformative priors or when influential observations are present in the data. [7, 8]

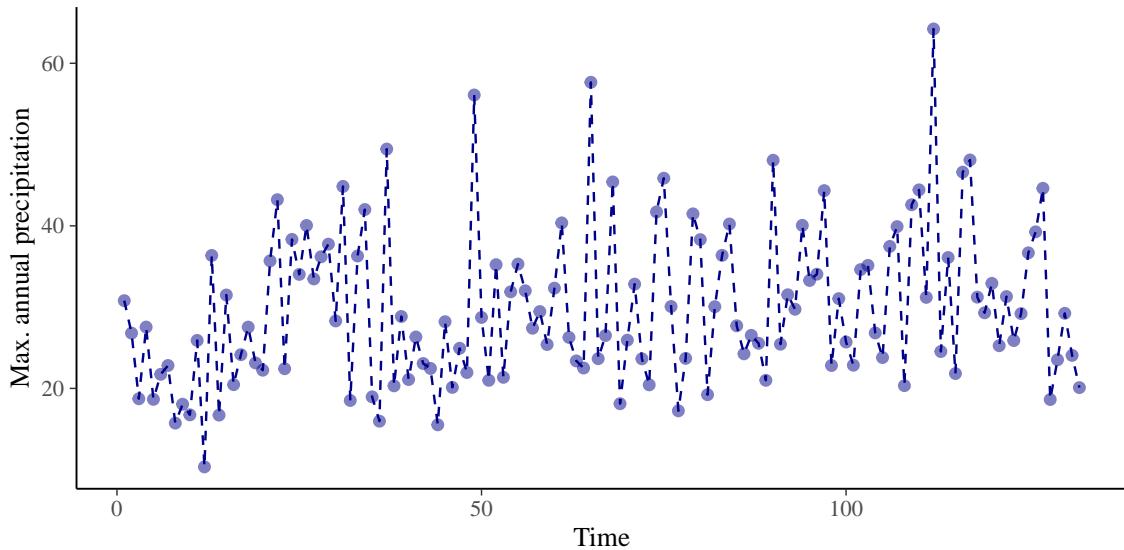
Samples from the posterior predictive distribution (PPD) are useful tool for diagnosing lack of fit. In order to obtain such samples, the parameter vectors  $\theta^{(1)}, \dots, \theta^{(N_{\text{iter}})}$  are used to generate samples  $y_{\text{rep}}^{(1)}, \dots, y_{\text{rep}}^{(N_{\text{iter}})}$  such that  $y_{\text{rep}}^{(m)} \sim \pi(y | \theta^{(m)})$ . The samples are standardised and used in posterior predictive checks, where summary statistics of the standardised data is compared to the distribution formed by the PPD samples with respect to said statistic. If the model fits the data well, one would expect the distribution to be close to the observed statistic.

## 2.3 Overview of data

The raw data originate from a weather station in the south of England and consist of 48212 measurements of cumulative precipitation in millimetres over a 24 hour period, from 12:00 to 12:00 the day after, spanning approximately 132 years. Tail observations are removed so that the data are exactly over a  $N = 132$  year period, where the measurements in year  $n$  will be denoted by  $X_1^{(n)}, \dots, X_{365}^{(n)}$  with  $n \in \{1, \dots, 365\}$ . The maximum annual precipitation is then defined by  $y^{(n)} = \max\{X_1^{(n)}, \dots, X_{365}^{(n)}\}$ . This is used to define the data vector  $y = \{y^{(1)}, \dots, y^{(N)}\}$  which constitutes the data used in the study. Table 1 shows summary statistics for  $y$ .

**Tab. 1.** Summary statistics of the aggregated data,  $y$ .

Min.	$Q_1$	Median	Mean	$Q_3$	Max.
10.35	22.82	27.95	29.88	36.12	64.22



**Fig. 1.** Line plot of the aggregated data,  $y$ .

## 3 Bayesian modelling framework

The simplest model is a fixed-parameter model, such that the response is assigned  $\text{GEV}(\mu, \sigma, \xi)$  likelihood where  $\mu \sim N(0, \sigma_\mu)$ ,  $\sigma \sim \text{Exp}(\lambda_\sigma)$ , and  $\xi$  is assigned a  $\text{Beta}(\alpha_\xi, \beta_\xi)$  prior which is translated onto the  $[-0.5, 0.5]$  interval. Furthermore,  $\sigma_\mu = 10$ ,  $\lambda_\sigma = 3$ , and  $\alpha_\xi = \beta_\xi = 4$  so that the parameter  $\xi$  has most density close to the origin, favouring small values of  $\xi$ , in the spirit of [9]. For non-stationary modelling, two hierarchical models, in which the location parameter  $\mu$  evolves with time, are presented. In the former such model,  $\mu_t = \mu_0(1 + \Delta(t - t_c))$ , where  $\mu_0 \sim N(0, \sigma_{\mu_0})$ ,  $\Delta \sim N(0, \sigma_\Delta)$  and  $\sigma_{\mu_0} = \sigma_\Delta = 10$ , and  $t_c = 66$  is the midpoint of elapsed time in the data set.

In the latter model, the location parameter  $\mu_t$  at time  $t$  is modelled using a piecewise linear trend  $\mu_t = \mu_0 + \sum_{j=1}^m \beta_j \mathbf{1}_{\{t>t_j\}}(t - t_j)$  where  $1 = t_1 < \dots < t_m < n$  is an equally-spaced partition of  $\{1, \dots, n\}$  into  $m$  timepoints. As before,  $\sigma \sim \text{Exp}(\lambda_\sigma)$ ,  $\mu_0 \sim N(0, \sigma_{\mu_0})$ , and  $\xi \sim \text{Beta}(\alpha_\xi, \beta_\xi)$  shifted onto the  $[-0.5, 0.5]$  interval. Three prior structures for the trend coefficients  $\beta_1, \dots, \beta_m$  are considered:

- (a) an i.i.d. prior where  $\beta_1 \sim N(0, \sigma_\beta^{(1)})$  and  $\beta_j \sim N(0, \sigma_\beta)$  for  $j \geq 2$ ,
- (b) a random walk prior  $\beta_1 \sim N(0, \sigma_\beta^{(1)})$  and  $\beta_j \sim N(\beta_{j-1}, \sigma_\beta)$  for  $j \geq 2$ , and
- (c) an AR(1) prior where  $\beta_1 \sim N(0, \sigma_\beta^{(1)})$ , and  $\beta_j \sim N(\phi\beta_{j-1}, \sigma_\beta)$  for  $j \geq 2$ .

Here,  $\sigma_\beta^{(1)} = 0.10$  while  $\sigma_\beta = 0.01$  to allow the trend to adapt to greater variations when the data exhibits a rapid rise or fall over time. In the AR(1) prior,  $\phi$  is assigned a  $\text{Unf}(-1, 1)$  prior to ensure stationarity. For each prior structure, the number of coefficients  $m$  is varied and optimised over WAIC and LOOIC to select the optimal architecture for the latent layer.

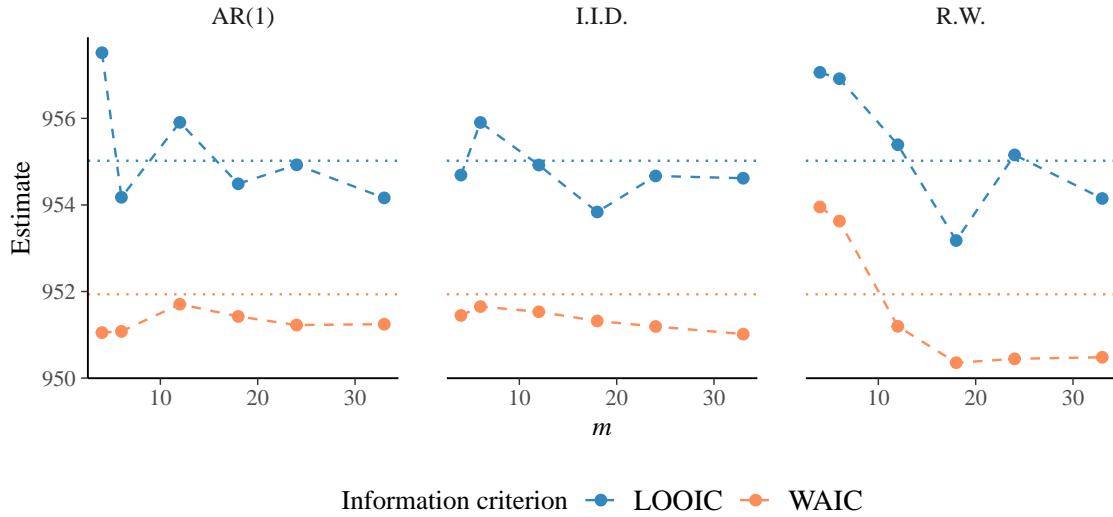
## 4 Results

All of the fitted models converged to their respective stationary distributions with  $\hat{R} \approx 1$  and  $\widehat{\text{ESS}}$  exceeding a hundred for all parameters. Figures 6-11 in Appendix A show trace and autocorrelation for the models, with the random walk prior representing the piecewise model. Noise-like behaviour observed in the trace plots is indicative of convergence, and a steep drop in the autocorrelation function  $\rho(l)$  for  $l > 5$  indicates independent sampling.

Table 2 shows LOOIC and WAIC estimates for each of the proposed models along with the optimal value of  $m$ . The piecewise model with the random walk prior resulted in the most favourable results with respect to WAIC and LOOIC despite relatively small differences in performance. However, the non-stationary models display a marked increase in accuracy compared to the stationary fixed-parameter model ( $\widehat{\text{LOOIC}} \approx 965$ ,  $\widehat{\text{WAIC}} \approx 962$ ). A line plot of WAIC and LOOIC estimates for the proposed models are shown in Figure 2. The optimal value of the hyperparameter  $m$  in the piecewise linear trend model is  $m = 18$ . In light of these results, the remainder of the model evaluation will focus on the random walk piecewise model with said value of  $m$ . Summary statistics for parameters  $\mu_0$ ,  $\sigma$ , and  $\xi$  is shown in Table 3, and an index plot of the  $\beta_j$  with 95% credible sets is shown in Figure 6.

**Tab. 2.** LOOIC and WAIC estimates for the proposed models. For the piecewise linear trend model, estimates for all three prior structures are shown for the optimal choice of hyperparameter  $m$ , shown enclosed in parentheses alongside information criteria estimates.

	Piecewise trend				
	Stationary	Linear trend	AR(1)	I.I.D.	R.W.
$\widehat{\text{LOOIC}}$	965.2	955.0	954.2 (6)	953.8 (18)	953.2 (18)
$\widehat{\text{WAIC}}$	961.9	951.9	951.1 (4)	951.0 (33)	950.3 (18)



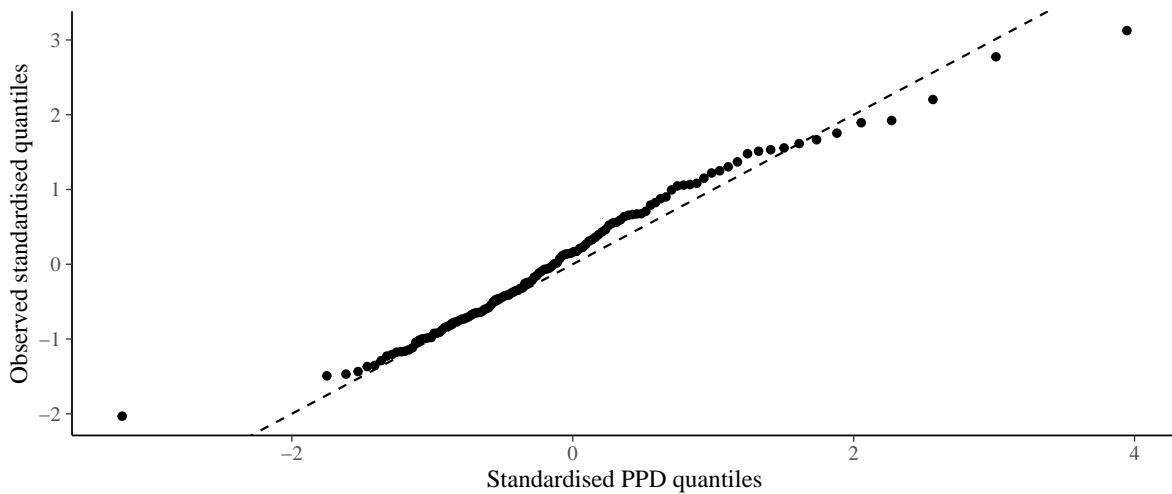
**Fig. 2.** Line plot showing WAIC and LOOIC estimates as a function of number of coefficients  $m$  in the piecewise linear trend model. Dotted lines show WAIC and LOOIC for the linear trend model.

The comparison of the quantiles of the standardised observed data with standardised PPD samples, shown in Figure 3, along with comparisons of empirical density function estimates, shown in Figure 4, indicate that the random walk piecewise model is a good fit for the data. Posterior predictive checks shown in Figure 5 with Bayes  $p$ -values displayed with each summary statistic show that the model is able to adapt to the data reasonably well to the data, with the maximum, mean and median values observed in the data falling within the 95% credible sets of the transformed PPD distribution. The minimum deviates the most, with a Bayes  $p$ -value of 0.14, which is far from being indicative of a lack of fit, although it suggests that the model slightly overestimates the minimum in the data. A comparison of density function estimates of a subset of samples from the standardised PPD, shown in Figure 4 further reinforces the aforementioned result, showing that the curves align closely.

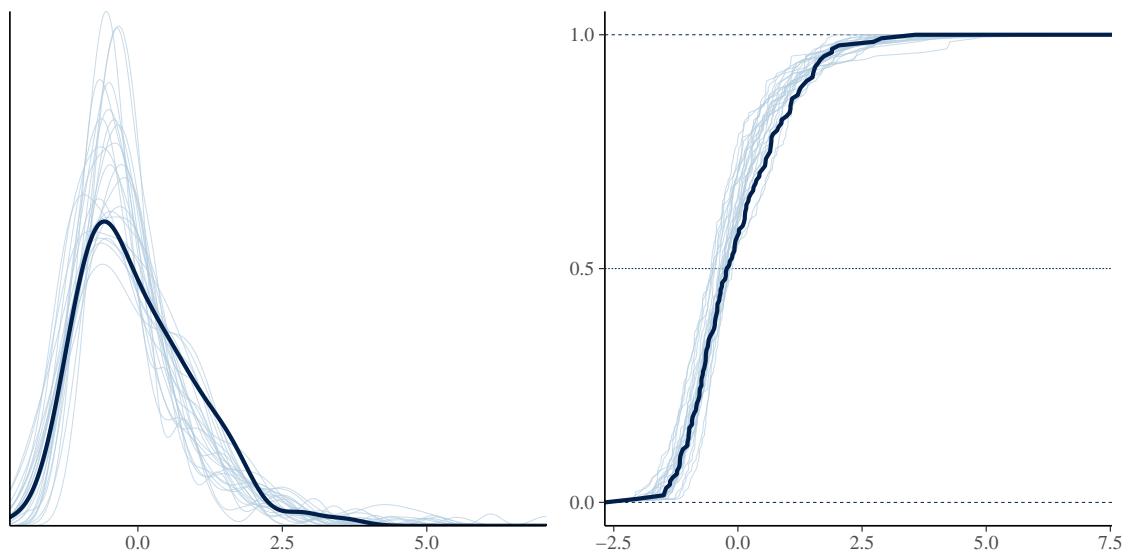
That the random walk prior structure is most effective in modelling the trend may be explained by the smoothness condition that is imposed on the  $\beta_j$  compared to the i.i.d. prior, where the coefficients are allowed to explore the sample space more freely. However, it seems that adding an autoregressive coefficient as in the AR(1) is detrimental in terms of accuracy, which could indicate that the added hyperparameter may be overfitting the data.

**Tab. 3.** Summary statistics for fitted parameters in the random walk piecewise model. The quantities  $q_{0.025}$  and  $q_{0.975}$  are the 2.5% and 97.5% quantiles, respectively. Thus, the 95% credible sets for the parameters are given by  $(q_{0.025}, q_{0.975})$ .

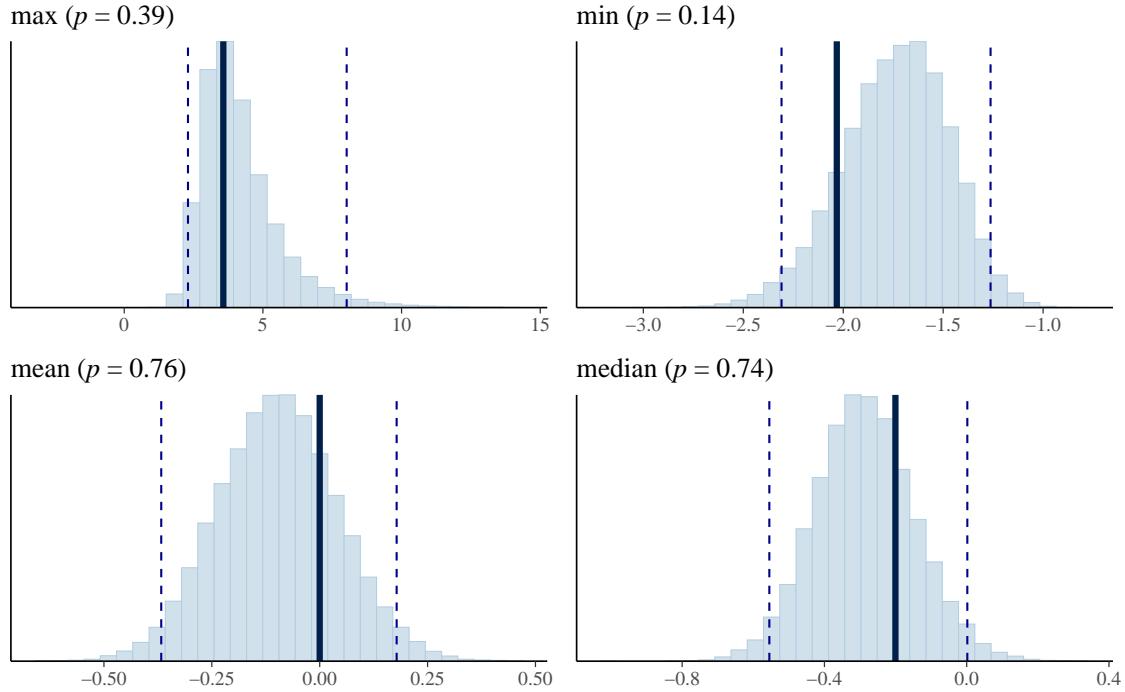
	Mean	Median	SD	$q_{0.025}$	$q_{0.975}$
$\mu_0$	20.7	20.7	1.34	18.0	23.3
$\sigma$	6.56	6.54	0.454	5.72	7.50
$\xi$	0.0593	0.0574	0.0629	-0.0587	0.187



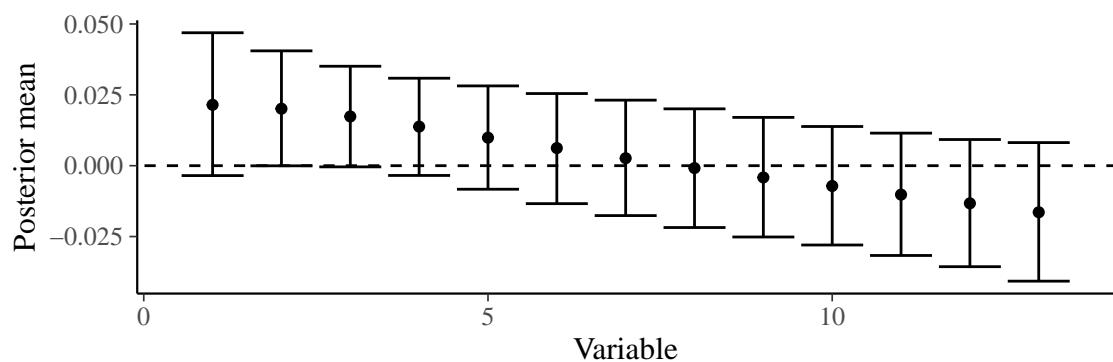
**Fig. 3.** Q-Q plot of standardised observed data against standardised PPD samples.



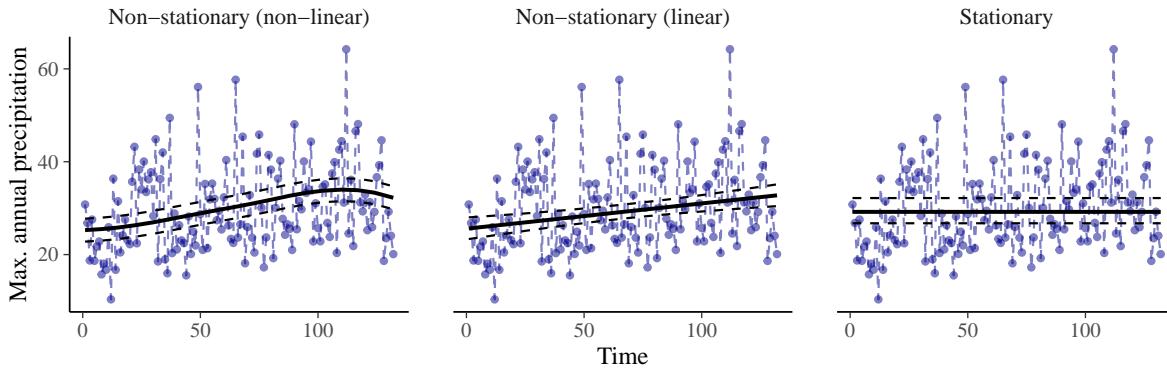
**Fig. 4.** (Left) Densities of a subset of standardised PPD samples shown in light blue in comparison to density estimate of standardised observed data, shown in darker blue. (Right) Empirical cumulative density function of a subset of standardised PPD samples in light blue compared to the ECDF of standardised observed data  $y$  in dark blue.



**Fig. 5.** Posterior predictive checks for the random walk piecewise model. The histogram shows estimates from the standardised posterior predictive sample matrix, and the vertical blue line represents the observed estimate from the data. Dashed lines show 95% probability bands. The quantity  $p$  represents the cumulative density evaluated at the estimated statistic, called the Bayes  $p$ -value.



**Fig. 6.** Index plot of coefficients  $\beta_1, \dots, \beta_m$ , showing their posterior mean with 95% credible sets.



**Fig. 7.** Plot comparing posterior mean of GEV distribution with location parameter  $\mu_t = \tau(t, \theta)$  with line plot of data. Here,  $\theta$  is a parameter vector modelled in the latent layer. Dashed lines indicate 95% probability bands.

## 5 Discussion

Model comparison using information criteria clearly favoured non-stationary models with a temporally evolving location parameter. Although differences in accuracy between variants of the piecewise model and the linear time trend model are small, the random walk model is most favourable when considering the LOOIC and WAIC information criteria. Empirical model evaluation using posterior predictive sampling showed that the model describes variability within the data reasonably well, with quantiles generated from the posterior predictive distribution and the data falling approximately on the identity line on a Q-Q plot, indicating the absence of lack of fit for the optimal model.

Improved accuracy in models that implement an evolving location parameter over time suggests that the mean annual maximum precipitation is variable with time. A line plot of the posterior mean of the GEV variable modelled in the selected piecewise model shows an upward trend from  $t = 0$  up to approximately one hundred years, followed by what seems to be a downward trend in the remaining thirty or so years of the data. Whether this trend will continue in this fashion remains to be seen. However, it is important to note that the trend observed in these data may not reflect what one would observe in other locations.

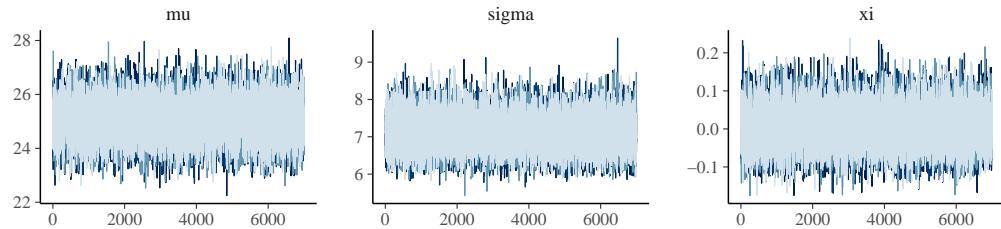
The modelling of extremes such as those explored in these data is an interesting topic in statistics, both in theory and practice. Limiting behaviours of block maxima ensured by the Fisher-Tippett theorem make the GEV distribution an easy candidate for the likelihood of data, but as shown in this article, the design of the latent layer in a Bayesian setting can vastly improve the accuracy of these models, such as in an iterated setting with parameters evolving over time. Additionally, the addition of spacial effects and the inclusion of covariates relevant to the measuring stations from which the data are sources can improve accuracy even further, e.g. in [1], [5], and [9]. By leveraging these advanced techniques, more precise estimates and better predictions of extreme events can be achieved.

## References

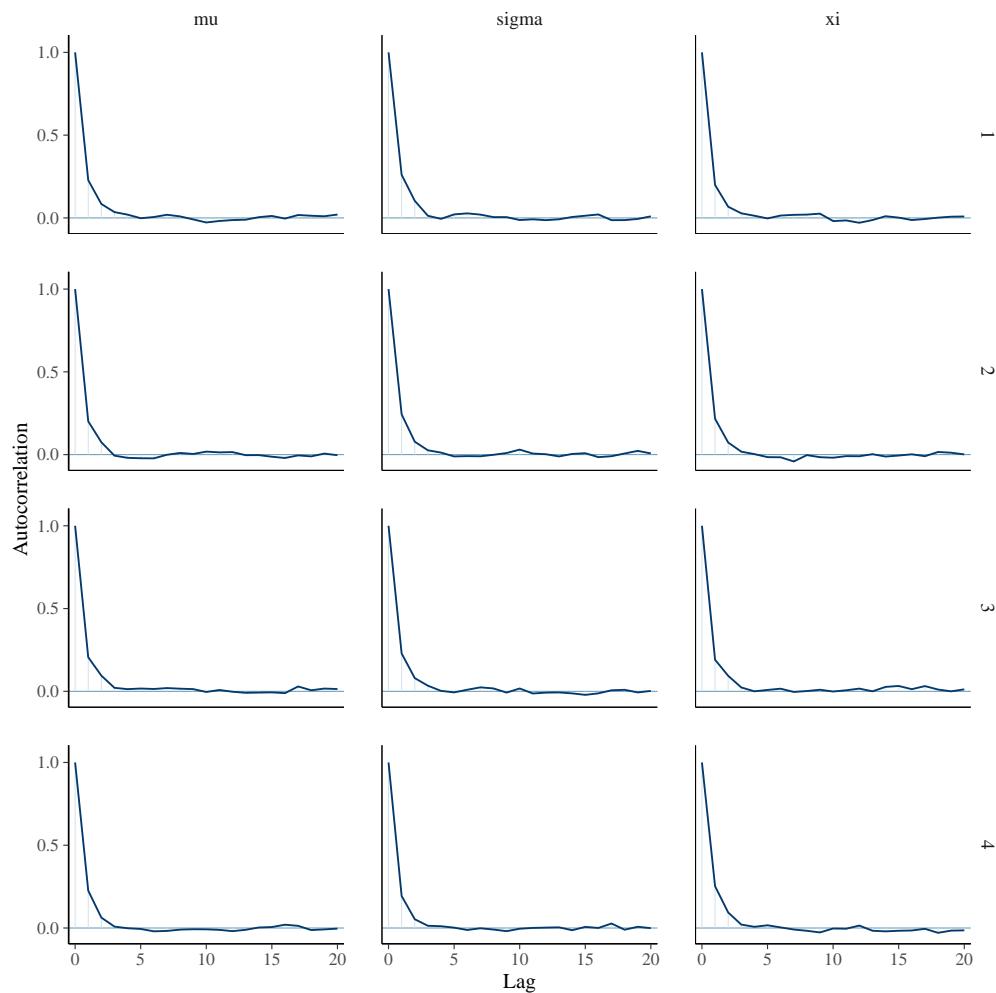
- [1] Anita Verpe Dyrrdal et al. ‘Bayesian hierarchical modeling of extreme hourly precipitation in Norway’. In: *Environmetrics* 26.2 (2015), pp. 89–106. ISSN: 1099-095X. doi: 10.1002/env.2301. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1002/env.2301> (visited on 16/01/2025).
- [2] Bob Carpenter et al. ‘Stan: A Probabilistic Programming Language’. In: *Journal of Statistical Software* 76 (Jan. 2017), pp. 1–32. ISSN: 1548-7660. doi: 10.18637/jss.v076.i01. URL: <https://doi.org/10.18637/jss.v076.i01> (visited on 16/01/2025).
- [3] R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing, 2024. URL: <https://www.R-project.org/>.
- [4] Stuart Coles. *An Introduction to Statistical Modeling of Extreme Values*. Springer Series in Statistics. London: Springer, 2001. ISBN: 978-1-84996-874-4 978-1-4471-3675-0. doi: 10.1007/978-1-4471-3675-0. URL: <http://link.springer.com/10.1007/978-1-4471-3675-0> (visited on 16/01/2025).
- [5] Arnab Hazra, Raphaël Huser and Árni V. Jóhannesson. ‘Bayesian Latent Gaussian Models for High-Dimensional Spatial Extremes’. In: *Statistical Modeling Using Bayesian Latent Gaussian Models : With Applications in Geophysics and Environmental Sciences*. Ed. by Birgir Hrafnelsson. Cham: Springer International Publishing, 2023, pp. 219–251. ISBN: 978-3-031-39791-2. doi: 10.1007/978-3-031-39791-2\_7. URL: [https://doi.org/10.1007/978-3-031-39791-2\\_7](https://doi.org/10.1007/978-3-031-39791-2_7) (visited on 16/01/2025).
- [6] Stan Development Team. *Convergence and efficiency diagnostics for Markov Chains — Rhat*. URL: <https://mc-stan.org/rstan/reference/Rhat.html> (visited on 16/01/2025).
- [7] Sumio Watanabe. ‘Higher Order Equivalence of Bayes Cross Validation and WAIC’. In: *Information Geometry and Its Applications*. Ed. by Nihat Ay, Paolo Gibilisco and František Matúš. Cham: Springer International Publishing, 2018, pp. 47–73. ISBN: 978-3-319-97798-0. doi: 10.1007/978-3-319-97798-0\_3.
- [8] Aki Vehtari, Andrew Gelman and Jonah Gabry. ‘Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC’. In: *Statistics and Computing* 27.5 (Sept. 2017), pp. 1413–1432. ISSN: 1573-1375. doi: 10.1007/s11222-016-9696-4. URL: <https://doi.org/10.1007/s11222-016-9696-4> (visited on 17/01/2025).
- [9] Árni V. Jóhannesson et al. ‘Approximate Bayesian inference for analysis of spatiotemporal flood frequency data’. In: *The Annals of Applied Statistics* 16.2 (June 2022). Publisher: Institute of Mathematical Statistics, pp. 905–935. ISSN: 1932-6157, 1941-7330. doi: 10.1214/21-AOAS1525. URL: <https://projecteuclid.org/journals/annals-of-applied-statistics/volume-16/issue-2/Approximate-Bayesian-inference-for-analysis-of-spatiotemporal-flood-frequency-data/10.1214/21-AOAS1525.full> (visited on 16/01/2025).

## A Convergence of MCMC samplers

### A.1 Stationary model

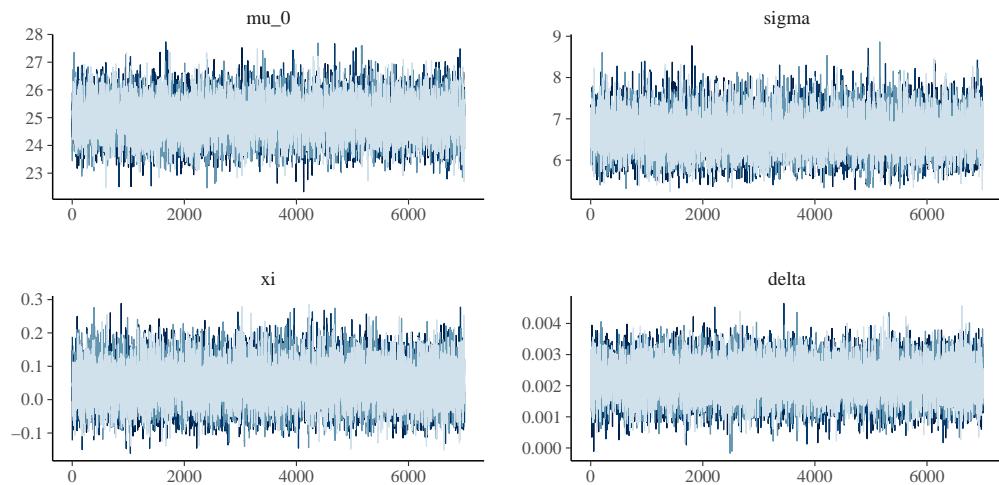


**Fig. 8.** Trace plot of the parameters for the stationary (fixed-parameter) model.

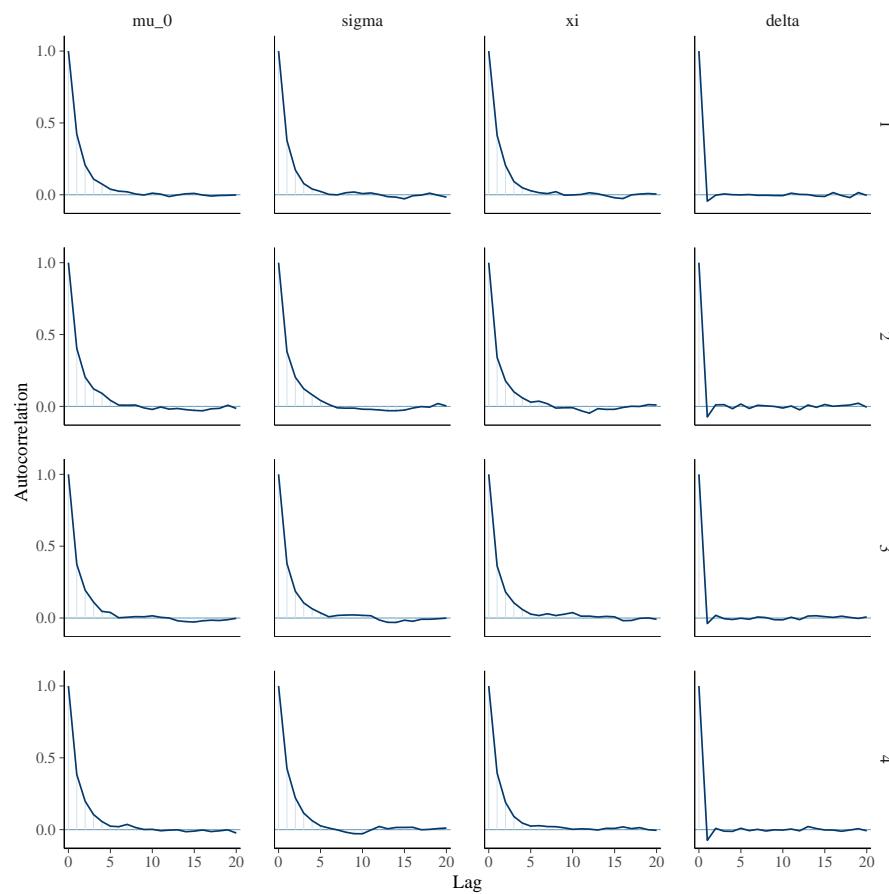


**Fig. 9.** Plot of autocorrelation function  $\rho(l)$  for the stationary model.

## A.2 Non-stationary model (linear trend)

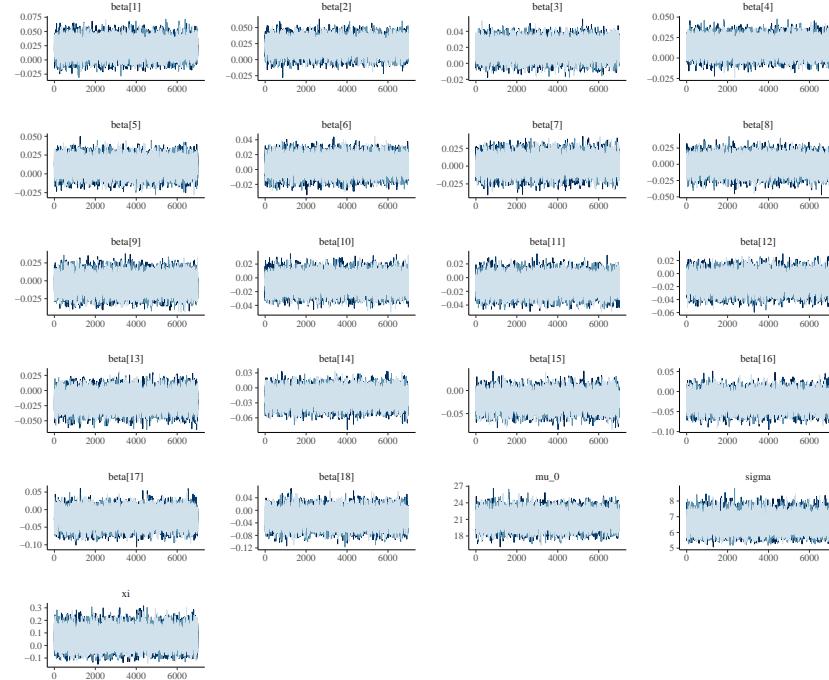


**Fig. 10.** Parameter summary statistics for the timetrend model.

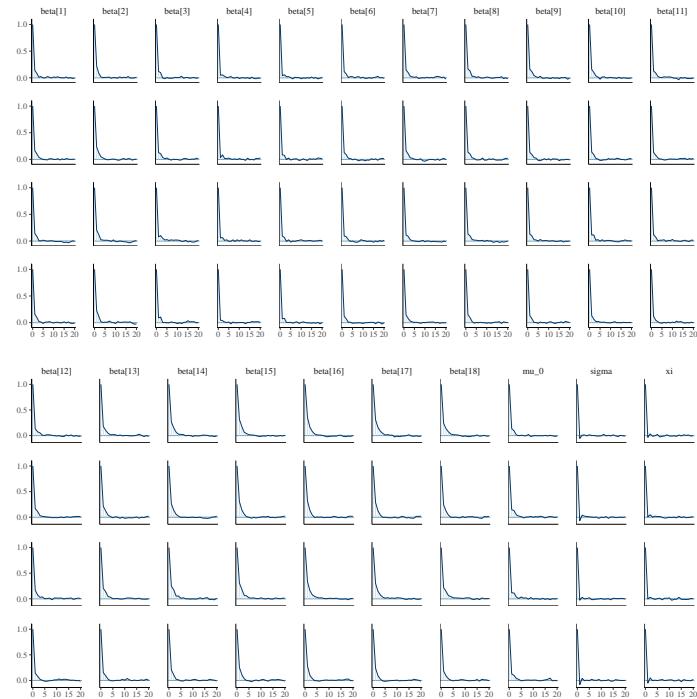


**Fig. 11.** Plot of autocorrelation function  $\rho(l)$  for the linear time trend model.

### A.3 Non-stationary model (non-linear, random walk prior)



**Fig. 12.** Trace plot of parameters of random walk prior piecewise model.



**Fig. 13.** Plot of autocorrelation function  $\rho(l)$  for the random walk prior piecewise model.